Volume 309, number 3, 311–315 FEBS 11522 September 1992

Cloning and sequence analysis of a cDNA clone from *Arabidopsis thaliana* homologous to a proteasome α subunit from *Drosophila**

Pascal Genschik, Gabriel Philipps, Claude Gigot and Jacqueline Fleck

Institut de Biologie Moléculaire des Plantes du CNRS, Université Louis Pasteur, 12 rue du Général Zimmer, 67084 Strasbourg Cedex, France

Received 8 July 1992

A cDNA clone isolated from an *Arabidopsis thaliana* cell suspension culture library showed considerable similarities to the proteasome 28 kDa α subunit of *Drosophila* [(1990) Gene 90, 235–241]. The 250 amino acid-long protein encoded by *Arabidopsis* TAS-g64 clone has important homologies in its primary structure and in the predicted secondary structure with the PROS-28.1 clone from *Drosophila*. The only divergence observed between the two sequences is for the 20 C-terminal amino acids. This subunit might share important functions in both kingdoms, as revealed by the important conservation between plants and animals. In plant cells it is encoded by a single-copy gene and probably regulated by stress and/or division.

Plant; *Arabidopsis thaliana*; Stress/division; Cell suspension culture; cDNA cloning; Proteasome

1. INTRODUCTION

Proteasomes are multicatalytic complexes consisting of a set of non-identical polypeptides, present in both nuclei and cytoplasm of a variety of eukaryotic cells from yeast to man [1–3] and also described in archaebacteria [4]. In eukaryotes these particles of a sedimentation coefficient of about 20 S have a molecular mass around 700 kDa and constitute about 15–20 different subunits ranging from 21 to 32 kDa [5–6], but with different pI values. These structures are strongly conserved over large evolutionary distances and play fundamental roles in the physiology of the cell. They have been reported to be involved in mRNA transcription [7] and tRNA processing [8], as well as tRNA degradation [9]. This 700 kDa particle is essential in non-lysosomal protein degradative processes. The 20 S proteasome is associated with other protein components to form a 26 S complex. This complex is involved in the ATP-dependent degradation of ubiquitin-conjugated proteins [10–16].

The primary structure of several of the proteasome subunits has been reported in a large variety of animal and yeast species (for references see [17]). At the structural level proteasomal bodies have been found in peas [18] and analysed in detail in wheat [1], but to our knowledge, their sequence has not been so far reported in higher plants.

In this paper we report and analyse the sequence of a full-length cDNA of a polypeptide of *Arabidopsis thaliana* that shows strong similarity with a 20 S proteasome α subunit (PROS-28.1) from *Drosophila* [19].

2. MATERIALS AND METHODS

The cDNA clone was isolated from a library made from a cell suspension culture of *Arabidopsis thaliana* 16 h after subculturing. This library was constructed in the expression vector lZAP by M. Axelos and D. Trémousaygue (INRA, Toulouse). The DNA sequencing was carried out by the dideoxy chain-termination method of Sanger [20], by using 19-mer oligodeoxyribonucleotides synthesized in an Applied Biosystem DNA synthesizer as primers. Computer sequence analyses were performed with the UW-GCG facilities [21] offered by EMBL, Heidelberg. The hydropathy index was calculated by the algorithm of Kyte and Doolittle [22] over a window of 11 amino acids, and was plotted as a function of amino acid position. The secondary structure was established according to Posh and Dancy de Marcillac (unpublished) by a vectorial representation of proteins based on the amino acid distribution diagram of French and Robson [23]. Northern blot and Southern blot analyses were performed as described previously [24]. The probes were labeled by the multiprime DNA labeling method with [α-^{32}P]dCTP.

3. RESULTS AND DISCUSSION

3.1. *Primary structure of the TAS-g64 clone*

As a part of a research program (GDR-genome *Arabidopsis*) we isolated, from a library established from an *Arabidopsis thaliana* cell suspension culture harvested at the beginning of the log growth phase, a cDNA clone

Correspondence address: J. Fleck, Institut de Biologie Moléculaire des Plantes du CNRS, Université Louis Pasteur, 12 rue du Général Zimmer, 67084 Strasbourg Cedex, France.

*The nucleotide and amino acid sequences are in the EMBL, GenBank under the accession number: X66825.

that showed high similarities with a 28 kDa α subunit proteasome from *Drosophila*. Using the FASTA service, we found that the cDNA insert, TAS-g64, revealed a 56% homology to the nucleotide sequence of the recently published clone, PROS-28.1. The clone has been further analysed and the sequence of the insert determined by the strategy given in Fig. 1A. The full-length cDNA (Fig. 1B) constitutes 1,087 nucleotides terminated by a 22 nucleotide-long poly(A) tail. The 3' non-coding region constitutes 322 nucleotides devoid of any polyadenylation signal, which in plants [25] is dispensable. The open reading frame encodes a putative 250 amino acid peptide. From this deduced amino acid sequence the molecular mass could be estimated to be 27,320 Da and the pI 7.44.

Using the PROSITE computer program we could further identify several consensus sequences, such as a KKST pattern matching for phosphorylation by cAMP-dependent protein kinase at position 48, a RE-FLEKNY pattern matching a putative phosphorylation site by a tyrosine phosphokinase (TPK) at position 171, and a NATG pattern matching an Asn-glycosylation site at position 161. Glycosylation of the proteasome α subunits has been reported in Archeabacteria [17]. The different consensus sequences are boxed in Fig. 1B.

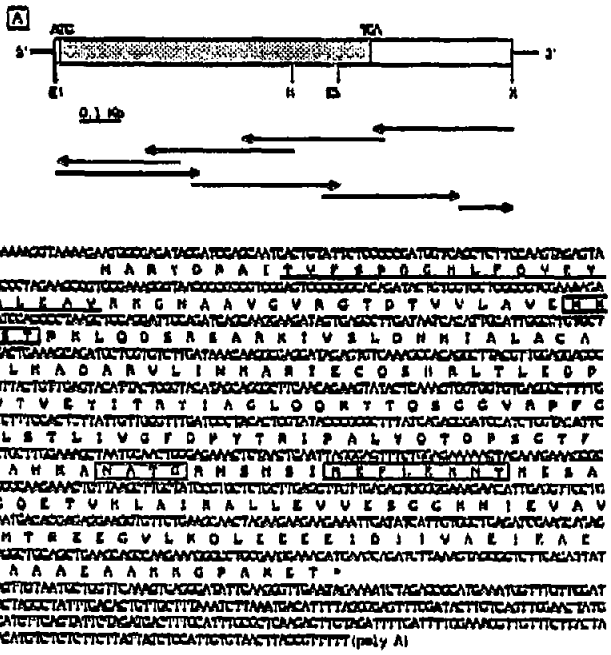

Fig. 1. Analysis of the cDNA clone, TAS-g64. (A) Restriction endonuclease map of the cloned TAS-g64 insert and sequencing strategy. The solid and open boxes indicate the coding and the 5'- and 3' non-coding regions, respectively. Continuous lines indicate the sequence of the vector. Sequenced regions are indicated by arrows. E1, *Eco*RI; E5, *Eco*RV; H, *Hin*dIII; X, *Xho*I. (B) Nucleotide sequence of the insert and deduced amino acid sequence are indicated below in the single letter code. Consensus sequences are boxed. The putative α subunit box [17] is underlined.

3.2. *Similarity to an α subunit proteasome of* Drosophila

Computer assisted homology analysis revealed that the highest identity of TAS-g64 putative encoded protein with any proteasome subunit so far reported is with PROS 28-1, an α subunit of *Drosophila*. The deduced amino acids of these two cDNA clones were compared by sequence alignment. Results are shown in Fig. 2A. The overall identity of the two protein sequences is 60%. Moreover, by analyzing the similarities of the amino acids as displayed by the UCWGCG GAP program, which takes into consideration the conservative changes, a similarity of 74% was reached. The identity between an animal and plant proteasome is even higher than between the two different α subunits reported in *Drosophila*, i.e. PROS-35 [26] and PROS-29 [27], where it is only around 30%. The calculated molecular weights of the two proteins of *Arabidopsis* and *Drosophila* are 27.3 and 27.9 kDa, respectively. The consensus sequences of the putative sites of Asn-glycosylation and Tyr-phosphorylation reported above are well conserved.

The similarity between the two proteins can be further demonstrated by the hydropathy profiles (Fig. 2B) established according to Kyte and Doolittle [22] and by the predicted secondary structure (Fig. 2C) calculated by Poch and de Marcillac (unpublished), on the basis of the amino acid distribution diagram estabished by French and Robson [23]. The two proteins have identical characteristics and are composed of two domains. The first domain covers the N-terminal region up to amino acid position 130. It is well conserved and contains the putative α type subunit box [17]. The second domain, covering the C-terminal region, is also well conserved but only up to amino acid position 230 where the two sequences diverged, becoming highly hydrophilic in *Drosophila*. It is noteworthy that this region contains a putative NTS (nuclear translocation signal) in *Drosophila* which is absent in *Arabidopsis*. This observation raises the problem of cellular localization. Nevertheless, all these observations suggest that these two proteins may constitute a subfamily of the α subunit proteasome, well conserved in the animal and plant kingdoms in which they probably share common functions.

3.3. *Genomic organization*

A DNA gel blot analysis was carried out to obtain information about the copy number of TAS-g64 encoding genes in the *Arabidopsis* genome. Genomic DNA was digested to completion with *Nco*I, an enzyme that has no restriction site inside the cDNA, and with *Eco*RV and *Hin*dIII, enzymes that have a site at position 693 and 577, respectively, (see Fig. 1) in the nucleotide sequence. Hybridization with the *Eco*RI–*Xho*I cDNA insert taken as a probe showed two major bands around, respectively, 9 and 4 kb with *Nco*I, and 0.8 and 1.8 kb with *Hin*dIII, but only one band at approxima-

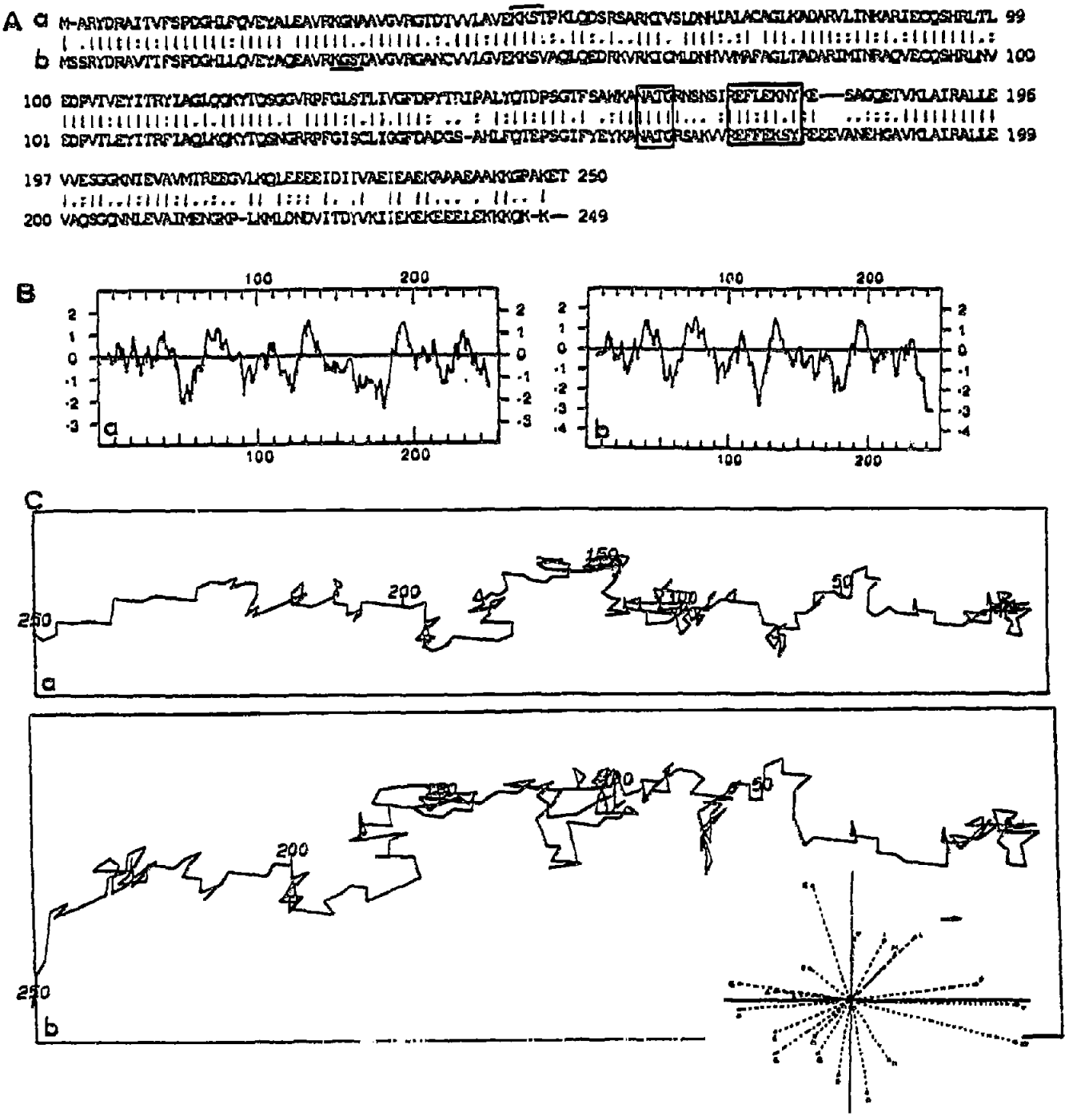

Fig. 2. Comparison of the predicted amino acid sequences of clone (a) TAS-g64 from *Arabidopsis* and (b) PROS-28.1 [19] from *Drosophila*. (A) Alignement of the two sequences. Gaps have been introduced to achieve maximum homology; vertical lines indicate amino acid identity, single and double dots indicate conservative changes, and dashes indicate gaps. Conserved consensus sequences are boxed, others are indicated by a line. (B) Hydrophobicity profile according to Kyte and Doolittle [22]. Negative values indicate hydrophilic regions. (C) Putative protein secondary structure according to Poch and de Murcillac (unpublished). The amino acid distribution diagram [23] is indicated.

tively 4 kb with *Eco*RV (Fig. 3). The presence of two bands on the *Nco*I blot may indicate the presence of a restriction site within an intron. The presence of only one band on the *Eco*RV blot could be explained by postulating that an intron disrupts this site in the gene. These results are consistent with the assumption of the existence of a single-copy gene. The minor hybridization signals found under high stringency hybridization

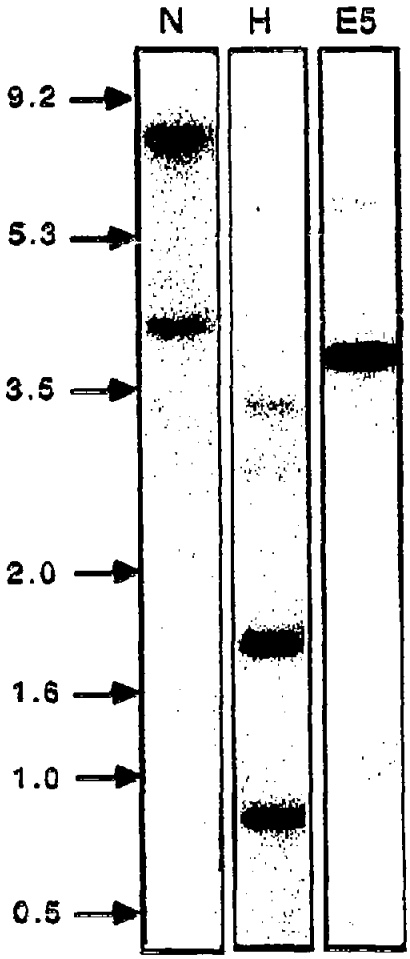

Fig. 3. Genomic Southern blot of *Arabidopsis thaliana* DNA digested with *Neo*I (N), *Hin*dIII (H) and *Eco*RV (E5), respectively, and hybridized with the *Eco*RI–*Xho*I TAS-g64 cDNA insert (Fig. 1). The position of size markers is indicated.

conditions indicate possible cross-hybridizations with the other proteasome subunits present in the *Arabidopsis* genome.

Analogous results have been obtained with genomic DNA blots from other plant species (results not shown), including tobacco, sunflower, wheat, oat, bean and red pepper, indicating that this gene is well conserved over species.

3.4. *Expression of the TAS-g64 gene*

In order to investigate how the TAS-g64 gene is regulated we analyzed the steady-state accumulation of TAS-g64-specific mRNA in several tissues and sliced leaf strips by Northern blot hybridization (Fig. 4). The integrity as well as the amount of mRNA had first been checked using a potato 25 S mRNA probe[28] and then hybridized to the *Eco*RI–*Xho*I insert (Fig. 1) used as a probe. Under stringent conditions we observed, as expected, one hybridization band at 1,100 nucleotides. At

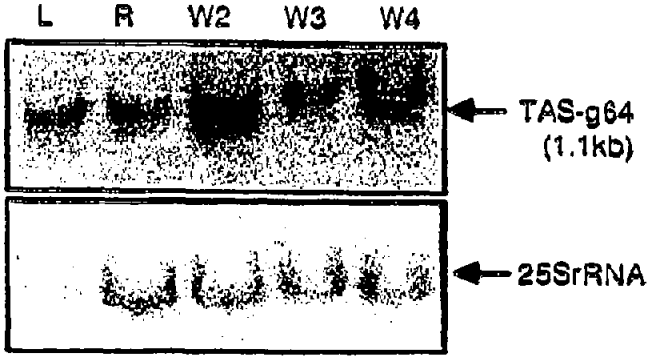

Fig. 4. RNA blot analysis of *Arabidopsis* TAS-g64 cDNA. 10 mg total RNA isolated from leaves (L), roots (R) and sliced leaf-strips harvested after 2 (W2), 3 (W3) and 4 (W4) days of incubation in the culture medium were hybridized with the *Eco*RI–*Xho*I TAS-g64 cDNA insert (Fig. 1) and with 25 S mRNA.

lower stringency and after extensive over-exposure of the gel (data not shown) a second band became faintly visible at 1,300 nucleotides, probably indicating cross-hybridizations with the subunits of the other proteasome families. The same level of accumulation of the 1.1 kb mRNA was observed in leaves (L), roots (R) and flowers (data not shown), while it significantly increased in sliced leaves. The signal was maximal in these sliced leaves two days after the incubation (W2) and then decreased. This observation seems to indicate that this mRNA accumulation is due mainly to the stresses caused by the slicing of the leaves. Nevertheless it is not possible at the moment to ascertain if this accumulation is only a stress response induced by the wound reaction, or if it has also a relationship with cell division. Indeed the conditions in which the leaf-strips have been incubated allow the very rapid re-initiation of mitotic activity, and important cell proliferation appears along the wounded edges. Moreover we showed recently that the two events, stress response and division, cannot be dissociated in the somatic cells during their re-entry into the cell cycle [29].

Acknowledgements: We wish to thank M.G. Daney de Marcillac for help in computer analysis and Mme A. Hoeft for the oligodeoxynucleotides. This work was supported by a grant from Ministère de la Recherche et de l'Espace to P.G. and by the GDR 1003ACS (CNRS).

REFERENCES

[1] Arrigo, A.P., Simon, M., Darlix, J.L. and Spahr, P.F. (1987) J. Mol. Evol. 25, 141–150.
[2] Falkenburg, P.E., Haass, C., Kloetzel, P.M., Niedel, B., Kopp, F., Kuehn, L. and Dahlmann, B. (1988) Nature 331, 190–192.
[3] Arrigo, A.P., Tanaka, K., Goldberg, A.L. and Welch, W.J. (1988) Nature 331, 192–194.
[4] Dahlmann, B., Kopp, F., Kuehn, L., Niedel, B., Pfeifer, G., Hegerl, R. and Baumeister, W. (1989) FEBS Lett. 251, 125–131.
[5] Orlowski, M. (1990) Biochemistry 29, 10289–10297.
[6] Rivett, A.J. (1989) Arch. Biochem. Biophys. 268, 1–8.
[7] Scherrer, K. (1990) Mol. Biol. Rep. 14, 1–9.

[8] Castano, J.G., Ornberg, R., Koster, J.G., Tobian, J.A. and Zasloff, M. (1986) Cell 46, 377–387.
[9] Tsukahara, T., Tanaka, K., Ogawa, T., Ishiura, S., Ryuhei, F. and Sugita, H. (1989) FEBS Lett., 255, 179–183.
[10] Matthews, W., Tanaka, K., Driscoll, J., Ichihara, A. and Goldberg, A.L. (1989) Proc. Natl. Acad. Sci. USA 86, 2597–2601.
[11] Eytan, E., Ganoth, D., Armon, T. and Hershko, A. (1989) Proc. Natl. Acad. Sci. USA 86, 7751–7755.
[12] Driscoll, J. and Goldberg, A.L. (1990) J. Biol. Chem. 265, 4789–4792.
[13] Heinemeyer, W., Kleinschmidt, J.A., Saidowsky, J., Escher, C. and Wolf, D.H. (1991) EMBO J. 10, 555–562.
[14] Orino, E., Tanaka, K., Tamura, T., Sone, S., Ogura, T. and Ichihara, A. (1991) FEBS Lett. 284, 206–210.
[15] Richter-Ruoff, B., Heinemeyer, W. and Wolf, D.H. (1992) FEBS Lett. 302, 192–196.
[16] Goldberg, A.L. and Rock, K.L. (1992) Nature 357, 375–379.
[17] Zwickl, P., Grziwa, A., Pühler, G., Dahlmann, B., Lottspeich, F. and Baumeister, W. (1992) Biochemistry 31, 964–972.
[18] Shelton, E., Kuff, E.L., Maxell, E.S. and Harrington, J.T. (1970) J. Cell Biol. 45, 1–8.
[19] Haass, C., Pesold-Hurt, B., Multhaup, G., Beyreuther, K. and Kloetzel, P.M. (1990) Gene 90, 235–241.
[20] Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
[21] Devreux, J., Haeberli, P. and Smithies, O. (1984) Nucleic Acids Res. 12, 387–396.
[22] Kyte, J. and Doolittle, R.F. (1982) J. Mol. Biol. 157, 105–132.
[23] French, S. and Robson B. (1983) J. Mol. Biol. 19, 171–175.
[24] Jamet, E., Durr, A., Parmentier, Y, Criqui, M.C. and Fleck, J. (1990) Cell Diff. Dev. 29, 37–46.
[25] Joshi, C.P. (1987) Nucleic Acids Res. 15, 9627–9640.
[26] Haass, C., Pesold-Hurt, B., Multhaup, G., Beyreuther, K. and Kloetzel, P.M. (1989) EMBO J. 8, 2373–2379.
[27] Haass, C., Pesold-Hurt, B. and Kloetzel, P.M. (1990) Nucleic Acids Res. 18, 4018.
[28] Logemann, J., Mayer, J.E., Schell, J. and Willmitzer, L. (1988) Proc. Natl. Acad. Sci. USA 85, 1136–1140.
[29] Criqui, M.C., Plesse B., Durr A., Marbach, J., Parmentier Y., Jamet E. and Fleck, J. (1992) Mech. Dev. (in press).